AI Doc Ruling Got Privilege Analysis Wrong

By **Matthew Coogan and Cindy Kuang** (March 26, 2026)

Editor's note: Law360 welcomes opinionated commentary and debate in our Expert Analysis section. To submit op-eds or rebuttals, or to speak to an editor about submissions, please email expertanalysis@law360.com.

Matthew Coogan

Cindy Kuang

On Feb. 10, in U.S. v. Heppner, U.S. District Judge Jed S. Rakoff of the U.S. District Court for the Southern District of New York held that roughly 31 documents reflecting a criminal defendant's interactions with a generative artificial intelligence platform were not protected by the attorney-client privilege or the work-product doctrine.

He acknowledged the ruling was "a question of first impression nationwide."

In a Feb. 17 memorandum in U.S. v. Heppner, the court framed the question in categorical terms: "[W]hen a user communicates with a publicly available AI platform in connection with a pending criminal investigation, are the AI user's communications protected by attorney-client privilege or the work-product doctrine? For the reasons that follow, the answer is no."[1]

The Heppner ruling is a watershed in the law governing use of generative AI to prepare for litigation. That makes the opinion's broad reasoning the more troubling: It treats use of an AI platform as dispositive disclosure to a third party, adopts an unduly narrow conception of work product and misreads Anthropic's consumer policies on which it relies, all while bypassing the existing doctrinal frameworks that should have governed the analysis.

The privilege, confidentiality and work-product questions AI presents are real — but existing doctrine already provides the tools to answer them. Heppner's error is not in confronting a novel problem, but in reaching for a categorical solution where none is warranted.

Background

In 2025, after learning through counsel that he was a target of a Southern District of New York criminal investigation, defendant Bradley Heppner — former CEO of Texas financial services firm Beneficient — used Claude, Anthropic's generative AI platform, to prepare written reports for discussions with counsel about his defense. According to the government's notes of calls with defense counsel at Quinn Emanuel Urquhart & Sullivan LLP, Heppner "would input facts into AI in order to provide response[s] to attorneys for legal advice" and did so "[f]or purpose[s] of discussing the issues with Quinn."[2]

Counsel represented that "every report" Heppner generated was "based on information counsel discussed with him in first instance" and was created as a "vehicle to consolidate [his] thoughts" for the "express purpose of talking to counsel."[3] While defense counsel represented that Heppner's reports were shared with and reviewed by counsel and facilitated their discussions, counsel acknowledged that they had not directed Heppner to generate the reports.

On Nov. 4, following a sealed indictment, Heppner was arrested and the FBI seized his devices containing the AI materials.[4] On Feb. 6, the government moved for a ruling that the materials were unprotected.[5] Four days later, at a previously scheduled conference on Feb. 10, without any written opposition submitted, the court granted the government's motion from the bench. The court rejected defense counsel's request to submit the AI materials for in camera review.[6] On Feb. 17, Judge Rakoff issued a memorandum explaining his bench ruling.

The strength of Heppner's privilege claim was not obvious. Counsel did not direct him to use Claude, and the record did not establish when the AI documents were created, what they contained or how closely they tracked subsequent attorney-client discussions. But that is precisely why a first-impression opinion demanded a more searching analysis than this one provided.

Heppner rests on the unstable premise that an AI is a third party.

The Heppner ruling repeatedly characterizes the defendant's use of Claude as "communication" with a "third party," a framing that drives the court's privilege, confidentiality and work-product analyses. The government's brief went further, asserting that "use of the AI tool here is no different than if [Heppner] had asked friends for their input on his legal situation."[7]

But disclosure to friends — human actors with independent judgment and freedom to repeat what they hear — is doctrinally different from submitting material to a software system. In Warner v. Gilbarco Inc., also decided on Feb. 10, the U.S. District Court for the Eastern District of Michigan captured the point more precisely: Generative AI systems are "tools, not persons, even if they may have administrators somewhere in the background."[8]

On attorney-client privilege, Heppner resolved the first element of the U.S. Court of Appeals for the Second Circuit's 2011 test in U.S. v. Mejia[9] by observing that "Claude is not an attorney."[10] Under the Mejia ruling, however, the relevant relationship is between client and counsel, not between client and software. Heppner never claimed that Claude was acting as his counsel. That suggestion originated in the government's brief, which asserted that the "AI tool is obviously not an attorney."[11]

The question the court should have addressed was whether the defendant's use of a software tool to prepare materials intended to facilitate subsequent consultation with counsel defeated privilege.[12] But the opinion never analyzed that question, instead importing the proposition that privilege depends on a "trusting human relationship," language drawn not from the Mejia ruling or the broader privilege case law, but from a 2025 article published in the Harvard Journal of Law & Technology.[13]

The Heppner ruling's framing of Anthropic's platform as "publicly available" compounds the problem.[14] In context, "publicly available" ordinarily describes content accessible to anyone — a published document, an open database. The opinion applies it to the platform, meaning anyone can create an account. But the accessibility of a service does not determine the accessibility of user data submitted to it.

The Heppner ruling's conflation matters because it embeds the confidentiality conclusion in the question presented: If the platform is publicly available, protection seems foreclosed before the analysis begins.

The work-product analysis is Heppner's most significant gap.

On work product, the court first held that Rule 16(b)(2) of the Federal Rules of Criminal Procedure did not apply because the AI documents were seized by warrant rather than sought in criminal discovery.[15] But Rule 16(b)(2) also protects "reports, memoranda, or other documents made by the defendant, or the defendant's attorney or agent, during the case's investigation or defense."[16]

The court then turned to a narrower common-law formulation protecting materials prepared "by or at the behest of counsel," holding that Heppner's AI materials were not protected because he created them on his own.[17]

Rule 16(b)(2) covers documents "made by the defendant," the category the court had before it. The court may have been right that the rule does not directly govern materials seized by warrant. But setting it aside without explanation produces an unjustifiable asymmetry: Defense-preparation materials receive less protection when seized by warrant than when sought through ordinary criminal discovery.

The Fourth Amendment provides its own protections for warrant-seized materials, but those protections address the legality of the search, not whether the materials, once lawfully obtained, retain their protected character.

The opinion also treats work product too narrowly on its own terms. Work-product doctrine protects the integrity of adversarial preparation, not only materials the lawyer personally prepared, which is why its waiver standard is narrower than privilege's: In the Second Circuit, work product is waived only by disclosure to an adversary or in circumstances substantially likely to result in disclosure to an adversary.[18]

The Heppner ruling never engaged in that analysis. It moved from the proposition that Anthropic was a third party to the conclusion that protection was lost, without asking whether disclosure to Anthropic was disclosure to an adversary or to a conduit substantially likely to deliver the materials to one. Under normal waiver principles, a generic reservation of the right to comply with legal process does not make Anthropic a conduit to an adversary, any more than similar language in mainstream email or cloud service terms does.

That omission matters because Anthropic is a software company, not a litigation adversary. The Eastern District of Michigan's Warner ruling recognized as much, holding that a party's use of ChatGPT did not waive work-product protection under the Federal Rules of Civil Procedure because generative AI systems are "tools, not persons," and because a work-product waiver turns on disclosure to an adversary, not merely to any third party.[19]

The Warner ruling is not the last word, but it identifies the right doctrinal question — which Heppner largely bypassed.

Heppner's confidentiality analysis misreads Anthropic's actual user policies.

The Heppner ruling's confidentiality analysis rests on a material mischaracterization of Anthropic's policy language.

Both the court and the government relied on Anthropic's Feb. 19 consumer privacy policy to establish that Anthropic collects users' inputs and outputs, uses them to train its models, and may disclose them to governmental authorities.[20] But the policy actually states that Anthropic "will not use" inputs or outputs to train its models, subject to narrow exceptions

such as opt-in consent, user feedback and trust-and-safety review.[21]

Anthropic's contemporaneous terms of service make the same commitment.[22] Both the court and the government appear to have conflated the privacy policy's general description of how Anthropic trains its models on publicly available data with the specific terms governing user inputs and outputs.

Other Anthropic policy provisions reinforce the point: The terms of service provide that users retain their rights in inputs while Anthropic assigns any of its rights in outputs to the user,[23] and the notice on model training[24] states that Anthropic's models are "specifically trained to respect privacy."[25] None of this guarantees confidentiality. But the court gave the policy language — even as misread — far more analytical load than it could bear, particularly because the record did not establish which terms were operative when Heppner created the documents.[26]

The court's treatment of Anthropic's policies is not an isolated error. Having made platform policy language central to its confidentiality inquiry, the opinion never examined the relevant policy set with the precision that premise required. The same looseness appears in the opinion's reliance on the Southern District of New York's Jan. 5 order in In re: OpenAI Inc. Copyright Infringement Litigation,[27] where the cited order involved proportionality in mass third-party discovery over deidentified ChatGPT logs — not attorney-client confidentiality in a criminal case.[28]

Yet, the Heppner ruling imported that order's language into a different context, expanding the source's reference to ChatGPT to "[another publicly accessible AI platform]," grafting a characterization the source never adopted onto authority whose reasoning did not support it.[29] The result is self-reinforcing: The Heppner ruling cites outside authority for a proposition that only appears in the citation because the opinion put it there.

More fundamentally, Heppner's confidentiality reasoning lacks any limiting principle. The privacy policy language on which the opinion relies most heavily — that Anthropic may disclose information when required or permitted by law — is standard legal-process language found across mainstream cloud services, including consumer email and cloud storage providers.[30]

No court has held that the attorney-client privilege disappears when a client uses Gmail or any other cloud service whose terms reserve similar disclosure rights. Generic contractual reservation of legal-process compliance cannot itself resolve confidentiality.

Existing doctrine already answers the question Heppner posed.

The introduction of AI into litigation preparation does not justify collapsing distinct and well-trod doctrines into a single premise about communication with a publicly available platform. Existing law provides the necessary frameworks for analyzing privilege and work-product protection in this setting, and Heppner's failure lies in not using them.

On confidentiality, the U.S. Bankruptcy Court for the Southern District of New York's 2005 ruling in In re: Asia Global Crossing[31] supplies the framework that the Heppner ruling should have applied: a contextual, multifactor analysis for assessing whether an expectation of privacy in electronic communications is objectively reasonable. Whatever the correct answer on the Heppner case's facts, it should have rested on that functional inquiry, not on boilerplate platform terms.

New York's Civil Practice Law and Rules, Section 4548, which provides that a privileged communication does not lose its character "for the sole reason" that it is transmitted electronically,[32] reinforces the point and will directly govern state court and diversity cases where Heppner's reasoning may be cited.

On privilege, the Second Circuit's 2006 ruling in U.S. v. DeFonte vacated a denial of privilege where the Southern District of New York failed to conduct an individualized inquiry into whether a witness's journal entries, which recorded material for later discussion with counsel, fell within the privilege's scope.[33]

DeFonte alone would not resolve the question presented in the Heppner case, but it demonstrates that the privilege question required the kind of fact-sensitive, communication-centered inquiry that is absent from Heppner. The court never examined when the AI materials were created or whether their substance was communicated to counsel, and it declined counsel's offer of in camera review.

The opinion's Kovel dicta, which stems from the Second Circuit's 1961 ruling in U.S. v. Kovel,[34] compounds the anthropomorphization problem. The Heppner ruling suggests that if counsel had directed their client to use Claude, the tool "might arguably be said to have functioned in a manner akin to a highly trained professional," qualifying as a Kovel agent.[35] But the Kovel framework applies to persons — accountants, translators, consultants — not to software.

To treat a software tool as a potential Kovel agent is to anthropomorphize it, repeating the same error that runs through the opinion's third-party analysis. The better framing is simpler: When counsel directs a client to prepare materials for defense discussions, the resulting documents are prepared "at the behest of counsel," the very formulation Judge Rakoff applied. Whether the client uses a typewriter, a word processor or a generative AI platform to do so does not change the character of counsel-directed work.

That reframing identifies the practically important scenario the Heppner ruling leaves open: a represented party using an enterprise AI platform at counsel's direction, under contractual confidentiality protections, to prepare litigation materials. That case is doctrinally stronger on privilege, confidentiality and work product; existing doctrine should handle it without difficulty. The Heppner case may ultimately prove limited to its facts. But as written, the opinion's categorical language risks foreclosing that analysis before it begins.

Courts confronting similar questions about lawyers' and clients' use of generative AI in litigation preparation should do what Heppner did not: Start from the premise that an AI is a tool, not a third party, and resolve the case by applying long-standing doctrine to the facts before them.

Matthew G. Coogan is a partner and Cindy X. Kuang is an associate at Lankler Siffert & Wohl LLP.

[1] United States v. Heppner, No. 25 CR. 503 (JSR), 2026 WL 436479, at 2 (S.D.N.Y. Feb.

17, 2026).

[2] Rothman Decl. Ex. D (Dkt. 23-4) (notes of Jan. 21, 2026, call).

[3] Rothman Decl. Ex. E (Dkt. 23-5) (notes of Feb. 2, 2026, call).

[4] Heppner at 2-3; Tr. of Nov. 10, 2025, Arraignment (Dkt. 16).

[5] Gov't Mot. (Dkt. 22).

[6] Tr. of Feb. 10, 2026, Conference (Dkt. 30) at 3 ("I'm happy to submit the reports for your review in camera"; THE COURT: "No."), 6.

[7] Gov't Mot. (Dkt. 22) at 7.

[8] Warner v. Gilbarco, Inc., No. 2:24-cv-12333 (E.D. Mich. Feb. 10, 2026) (Dkt. 94) at 12.

[9] 655 F.3d 126, 132 (2d Cir. 2011).

[10] Heppner at 5.

[11] Gov't Mot. (Dkt. 22) at 8.

[12] See, e.g., United States v. DeFonte, 441 F.3d 92, 96 (2d Cir. 2006) (holding that "an outline of what a client wishes to discuss with counsel — and which is subsequently discussed with one's counsel — would seem to fit squarely within our understanding of the scope of the privilege"). The government's brief addressed DeFonte only in a footnote, distinguishing it by asserting, without evidentiary cite, that "[o]nly after [Heppner's] AI analysis was complete did the defendant share the AI output with his attorneys." Gov't Mot. (Dkt. 22) at 10 n.4. Whether the AI documents were created contemporaneously with or after attorney consultations was not established in the record.

[13] Heppner at 6 (citing Ira P. Robbins, Against an AI Privilege, JOLT Dig., Harvard L. Sch. (Nov. 7, 2025)). The language "trusting human relationship" appears in the cited blog post, not in Mejia or any other case cited in the opinion.

[14] Id. at 2.

[15] Id. at 9.

[16] Fed. R. Crim. P. 16(b)(2)(A).

[17] Heppner at 9-10.

[18] In re: Steinhardt Partners LP, 9 F.3d 230, 235 (2d Cir. 1993); see also United States v. Am. Tel. & Tel. Co., 642 F.2d 1285, 1299 (D.C. Cir. 1980) ("[W]hile the mere showing of a voluntary disclosure to a third person will generally suffice to show waiver of the attorney-client privilege, it should not suffice in itself for waiver of the work product privilege.").

[19] Warner, at 11-12.

[20] Gov't Mot. (Dkt. 22) at 9 ("Anthropic explicitly advises its users in its Privacy Policy ... that it uses this data to 'train' its AI tool."); Heppner at 6 (adopting the characterization that

Anthropic "uses such data to 'train' Claude"). Effective September 28, 2025, Anthropic's Privacy Policy was modified to, among other things, require users to choose whether to share their data with Anthropic's AI models. As Anthropic explained in an August 28, 2025, announcement of the change: "We're now giving users the choice to allow their data to be used to improve Claude and strengthen our safeguards against harmful usage like scams and abuse. Adjusting your preferences is easy and can be done at any time." See Anthropic, Updates to Consumer Terms and Privacy Policy, available at https://www.anthropic.com/news/updates-to-our-consumer-terms.

[21] Anthropic, Privacy Policy (eff. Feb. 19, 2025), § 2 ("We will not use your Inputs or Outputs to train our models, unless: (1) your conversations are flagged for Trust & Safety review ... , or (2) you've explicitly reported the materials to us ... , or (3) you've otherwise explicitly opted in to the use of your Inputs and Outputs for training purposes."). This policy was filed on the docket. See Dkt. 33.

[22] Anthropic, Terms of Service — Consumer (eff. Feb. 19, 2025), § 4 ("We will not train our models on any Materials that are not publicly available."). The identical language appears in the April 15 and May 1, 2025, versions of the TOS.

[23] Id. ("As between you and Anthropic, and to the extent permitted by applicable law, you retain any right, title, and interest that you have in the Inputs you submit. Subject to your compliance with our Terms, we assign to you all of our right, title, and interest — if any — in Outputs.").

[24] This was explicitly cross-referenced in the Privacy Policy that the court and the government relied on.

[25] Anthropic, Notice on Model Training (eff. Dec. 17, 2024), § 4 ("[O]ur models are specifically trained to respect privacy."). The Privacy Policy itself directs users to this document. See Privacy Policy, Introduction ("For information about how we collect and use personal data to develop our language models ... please see our Notice on Model Training.").

[26] Heppner at 6; Privilege Log (Dkt. 23-2) (providing no creation dates for AI documents).

[27] See Heppner at 6-7.

[28] In re: OpenAI Inc., Copyright Infringement Litig., No. 25 MD 3143, ECF No. 1021 (S.D.N.Y. Jan. 5, 2026). Judge Stein's actual language was comparative: "[P]rivacy interests in the wiretapped recordings of private phone conversations in [a government enforcement case] are stronger than the privacy interests in users' conversations with ChatGPT which users voluntarily disclosed to OpenAI." The court found users' privacy interests adequately safeguarded by de-identification, a protective order, and a reduced sample size — a proportionality finding in the context of aggregate, anonymized discovery that says nothing about an individual criminal defendant's confidentiality expectations.

[29] The government's brief used the bracket "[an AI tool]"; the court expanded it to "[another publicly accessible AI platform]."

[30] See, e.g., Google, Privacy Policy (providing that Google "may share" information "[f]or legal reasons," including "to meet any applicable law, regulation, legal process, or enforceable governmental request"). Google's Transparency Report documents compliance

with tens of thousands of government data requests annually.

[31] In re: Asia Global Crossing Ltd., 322 B.R. 247, 257 (Bankr. S.D.N.Y. 2005).

[32] N.Y. C.P.L.R. § 4548.

[33] DeFonte, 441 F.3d at 96.

[34] 296 F.2d 981 (2d Cir. 1961).

[35] Heppner at 7.